Preparative *In Vitro* Biosynthesis of Complex Polyketides

Ryan T. Riddle

The University of Texas at Austin

Department of Chemistry and Biochemistry

Undergraduate Class of 2010

An Honors Thesis for the

Dean's Scholars Honors Program

Submitted December 2009

Adrian T. Keatinge-Clay, Ph.D.

Chemistry and Biochemistry

Supervising Professor

John F. Stanton, Ph.D.

Chemistry and Biochemistry

Honors Advisor

Abstract:

My Dean's Scholars thesis project concerns the development of an enzymatic system for the *in vitro* production of polyketides, which we hope will lead to the discovery of new molecules and new medicines. Modular polyketide synthases (PKSs) are massive enzymes that act as molecular assembly lines to elegantly transform small carboxylic acids into complex polyketides. Many natural polyketides are currently used for their medicinal effects, including the antibiotic erythromycin, the immunosuppressant rapamycin, and the anticancer agent epothilone. A long-term goal of the Keatinge-Clay group is to use modified PKSs to create synthetic polyketides, which we hope will lead to new pharmaceutical drugs and valuable chiral intermediates for organic synthesis. Unfortunately, both the study and practical utility of PKSs are hampered by the low yield of their polyketide products – yields from the current protocol rarely exceed one milligram per liter of culture. To create a more efficient, reduced-component platform for studying engineered PKSs and to increase their yield, we devised a system for the preparative *in vitro* production of complex polyketides. This system includes enzymes necessary for the production of polyketide intermediates from propionate, a modified PKS to form a synthetic polyketide, and enzymes that aid in the regeneration of ATP and NADPH from polyphosphate and glucose, respectively. Currently, we have isolated and assayed most enzymes in the system and are optimizing the system conditions for polyketide yield. We hypothesize that gram quantities of complex polyketides will be produced using the *in vitro* system, greatly increasing the time and energy efficiency of polyketide synthesis.

I. Introduction:

For three and a half billion years, life on Earth and its chemistry have been evolving. Along the way, this process has bestowed upon us an innumerable amount of awe-inspiring compounds: DNA, proteins, polysaccharides, vitamins, and the other stuff of life. We humans are particularly interested in these biological molecules, as some of them – such as the essential nutrients – are necessary prerequisites for our lives, while others – like medicines and fibers – help to prolong and improve them. It follows that the primary purpose of biochemistry, the study of the chemistry of life, is to gain a better understanding of these molecules and the reactions that surround them so that we can use this knowledge to further improve and sustain life. In the following pages, I describe the development of a novel biochemical system that will help to further that purpose by allowing us to more easily and more efficiently produce new molecules and new medicines.

Of principal interest to the Keatinge-Clay group are polyketides, a few of which are shown in **Figure 1**. A wide array of bacteria, fungi, and plants synthesize these small yet

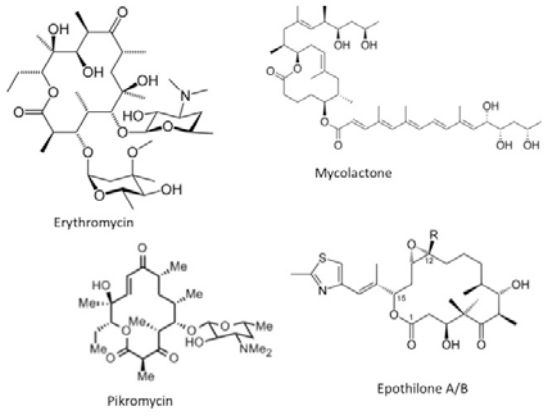

Figure 1. A few examples of polyketides

complex and diverse molecules (1). These organisms use polyketides for an extensive variety of purposes, including defense, information transfer, and pigmentation (2). Researchers have become interested in these molecules because many of the natural polyketides have been shown to have interesting medicinal effects, including antibiotic, anticancer, immunosuppressant, and cholesterol-lowering activity (3). While these properties are what make polyketides useful to humans, the reason they are so fascinating to researchers is that as complex and diverse as polyketides are, their build-

ing blocks are some of the simplest in all of biochemistry.

Polyketide synthases (PKSs) are nature's elegant method of transforming humble carboxylic acids into complex polyketides (**Figure 2** (4)). These impressive proteins are some of the largest enzymes found in all of nature. Type I modular PKSs in particular are composed of several distinct modules, each of which is a giant multi-domain enzyme itself. These PKSs work like molecular assembly lines, each module acting as a separate station, adding another extender unit to the growing chain and then modifying the resulting polyketide. Since each module can modify the growing chain in a different way, a great level of complexity can be obtained. The result is an extremely diverse group of natural products.

PKS modules exhibit behavior described as promiscuous, meaning that they can accept growing polyketide chains of varying lengths with remarkable tolerance (5,6). This allows researchers to rearrange PKS modules, forming enzymes that can produce novel, synthetic polyketides. By exploiting this property, it is theoretically possible to synthesize prodigious collections of unique and novel polyketides, and a long-term goal paramount to the Keatinge-Clay group is to produce exhaustive libraries of these synthetic polyketides (**Figure 3**).

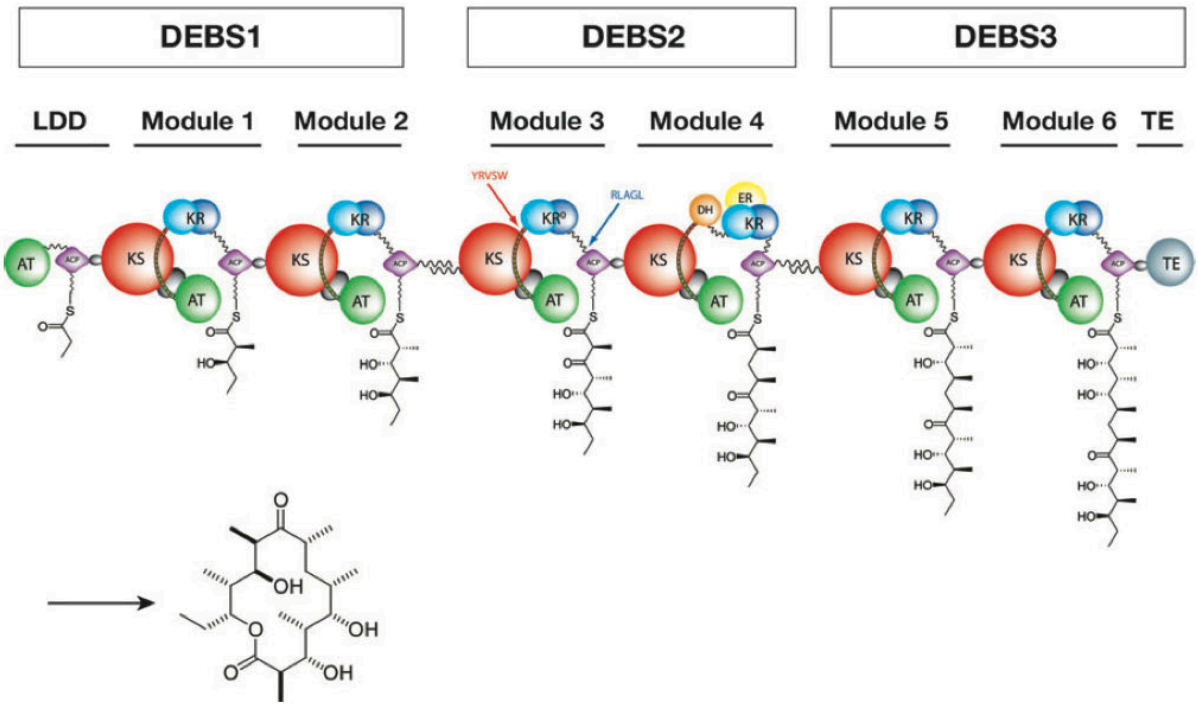

Figure 2. 6-DEB synthase transforms propionate into 6-DEB, a complex polyketide.

Many laboratories share the goal of producing large numbers of synthetic polyketides due to the vast number of new molecules and new medicines that could potentially be produced. Libraries of polyketides will serve as excellent leads for drug screening, which could lead to the next generation of pharmaceutical drugs. Researchers are especially interested in making drug-screening libraries out of polyketides because while many of the existing polyketides have useful medicinal effects, the natural polyketides represent only a tiny fraction of the molecules that could potentially be produced by rearranging PKS modules (7). In addition, polyketide libraries will serve as an unparalleled resource for or-

ganic chemists who will use polyketides as substrates for the synthesis of even more complex "unnatural" natural products. Polyketides are especially useful for this purpose due to their chirality, a property that is difficult for organic chemists to introduce into their creations by tradition methods (8).

It is no wonder, then, why many researchers are so interested in pursuing libraries of synthetic polyketides. Unfortunately, one obstacle standing in the way of the realization of this goal is the current protocol for polyketide production. One must grow and induce several liters of bacteria engineered to produce the particular polyketide and subsequently extract the supernatant in which they grow with large amounts of organic solvent. Emulsions form upon this extraction, and the solid obtained when this solvent is evaporated is extremely contaminated. One must then resuspend the sample and purify the polyketide with column chromatography. Finally, this laborious procedure results in a low yield of polyketide.

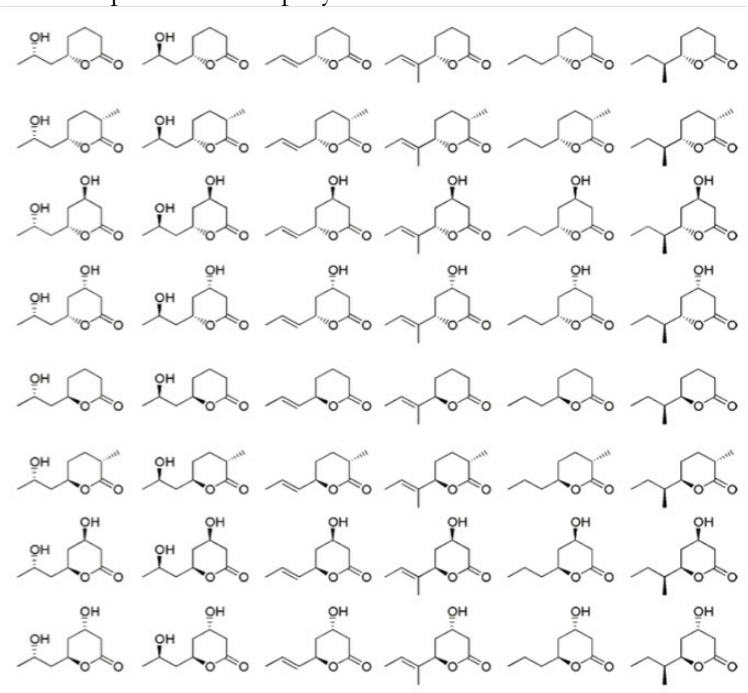

Figure 3. A sample polyketide library

We hope to improve the efficiency and yield of polyketide production dramatically by developing a novel system for the *in vitro* biosynthesis of polyketides. Whereas engineered bacteria have ulterior motives, including reproduction and metabolism of food, a reduced-component *in vitro* system would have the sole purpose of producing polyketides, resulting in the formation of fewer byproducts. Furthermore, the resulting system would be a black box of sorts in which a variety of purified PKSs could be inserted to produce pure polyketides. In order to construct such a system, it is necessary to identify and isolate those enzymes necessary for polyketide production. In addition, enzymatic systems for the regeneration of energy-carrying molecules must be included. Finally, all of the parameters in the system must be optimized to provide the best conditions for polyketide production. By constructing and optimizing such an *in vitro* system, we hypothesize that the yield of polyketides will be dramatically increased, making the dream of attaining polyketide libraries a much more feasible one.

The project of developing an *in vitro* system for the preparative biosynthesis of complex polyketides is an involved and arduous task. The gene for each enzyme in the system must be cloned and inserted into an expression

vector. Each enzyme must be purified and individually tested for activity with a unique assay. The ideal buffer conditions must be determined and implemented, and the concentrations of every component of the system must be optimized for polyketide yield. While the development of this system is a demanding and challenging project, its pursuit is ultimately worth the work it requires, for the result will be the potential for a generation of new molecules and new medicines.

II. Background:

A. DEBS: a Paradigmatic PKS

Erythromycin (**Figure 1**) is the quintessential polyketide. Ely Lilly discovered this powerful antibiotic in 1949 (9), and it remains in use today. Erythromycin was the first macrolide discovered, and its aglycone precursor, 6-deoxyerythronolide B (6-DEB) (**Figure 2**) remains the most studied product of any PKS. The antibiotic factory that assembles 6-DEB, DEB synthase (DEBS), serves as the paradigmatic PKS. Amazingly, the six modules of DEBS utilize little more than seven molecules of propionate and a few cofactors to manufacture the complexity of 6-DEB. The study of the structure and function of DEBS has helped to elucidate the mechanisms by which PKSs function to create such complex and diverse products.

1. Fatty Acid and Polyketide Biosynthesis

Each module of a PKS like DEBS is analogous to a fatty acid synthase (FAS), with a few modifications. In fatty acid synthesis, the FAS accepts sequential acetate units, adding another two carbons to the growing fatty acid with each additional acetate unit. The biggest difference between a FAS and a PKS module is that whereas the FAS must completely reduce each additional acetate unit, a PKS module can leave the carbonyl carbon of the extender unit in any of four reduction states (**Figure 4** (10)). The module can leave the carbonyl as is, reduce it to form a hydroxyl group, dehydrate the carbon to form a C-C double bond or completely saturate the growing chain just as the FAS would. Thus, we begin to understand why PKSs are responsible for the complexity and diversity of polyketides while FASs produce the relative simplicity and uniformity of fatty acids.

Type I modular PKSs like DEBS have another property that allows their products to be much more diverse than fatty acids. FASs are iterative, meaning that a FAS performs the same action on each additional extender unit until a particular chain length is reached. Modular PKSs, however, align multiple modules together in a molecular assembly line. Each module adds one additional extender unit to the growing chain and subsequently passes the developing polyketide to the next module. Since each module can reduce its extender unit to a different oxidation state, the product can contain many different functional groups in a unique order, which results in a much more interesting product than a fatty acid.

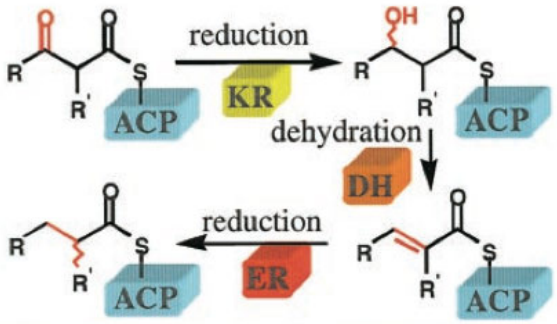

Figure 4: Fatty Acid/Polyketide Reduction. Whereas a fatty acid synthase must complete the complete loop shown, a PKS module may stop the reduction at any point.

Yet another property of many PKSs that facilitates complexity is their ability to produce chiral compounds. A PKS has three opportunities to introduce chirality to its product. The formation of a double bond allows for the production of either the *cis* or *trans* isomer of the polyketide. In addition, a stereocenter is formed when the carbonyl group of the extender unit is reduced to a hydroxyl group, and another arises with a peripheral methyl group if propionate is used instead of acetate, as in the case of DEBS. The power of PKSs to exhibit

stereoselectivity, oxidative control, and modularity enables PKSs to manufacture complex macrolides like 6-DEB.

From Propionate to Polyketide

From its life as a Lilliputian carboxylic acid to its incorporation into 6-DEB, propionate takes an epic journey. Before it can be used by DEBS, propionate must first be altered into methylmalonyl-CoA. In one route to this form, propionate-CoA ligase first joins propionate to coenzyme A, using ATP in the process. Next, propionyl-CoA carboxylase uses bicarbonate and another ATP to carboxylate propionyl-CoA, forming methylmalonyl-CoA. This activates the substrate to increase the thermodynamic favorability of polyketide chain extension. In addition, propionyl-CoA carboxylase exhibits stereocontrol, forming only the (2S) isomer of methylmalonyl-CoA, which is the only form that DEBS can accept. After these steps are performed, propionate is finally ready for its date with the polyketide assembly line.

DEBS is a prodigious protein, consisting of six modules on three peptides connected by interpolypeptide linkers which are then dimerized to form a massive protein complex (11). Each module of DEBS consists of several catalytic domains, similar to a FAS. Three of these domains are responsible for chain length extension, three reduce the extender units to varying degrees, and a terminal thioesterase cleaves the finished polyketide from DEBS upon completion of the assembly line.

The first of these domains, the acyl carrier protein (ACP) binds to methylmalonyl-CoA and guides it along the hardware of the module until all the chemistry is complete. The ACP is posttranslationally bound to a phosphopantethienyl arm via a conserved serine residue. This arm takes the place of coenzyme A on methylmalonyl-CoA as it bonds to the ACP, and its flexibility allows the ACP to shuttle the growing polyketide chain to each of the domains in the module.

The first action of the ACP is to travel to the acyl transferase domain (AT), where the AT catalyzes the binding of an extender unit to the ACP. For all six of the modules of DEBS, this extender unit is methylmalonate, but an additional AT joins an initial starter unit of propionate in the loading didomain (LDD) in front of module 1.

After it travels to the AT, the ACP next moves to the adjacent β-ketoacyl-S-ACP synthase domain (KS) where the growing polyketide chain is extended. The purpose of the KS is to accept the growing polyketide chain from the previous module (or LDD in the case of module 1) and catalyze chain extension. Just as in fatty acid synthesis, the decarboxylation of methylmalonate on the ACP allows the growing polyketide chain to be transferred from the KS to the extender unit on the ACP. The entire polyketide chain now extends from the phosphopantethienyl arm of the ACP and is ready to move to the reductive section of domains.

The first stop for the growing chain in the reductive loop of domains is the ketoreductase (KR). This domain reduces the carbonyl in the β position of the polyketide chain to form a hydroxyl group. In the process, a stereocenter is formed, and a particular KR will always catalyze the formation of a hydroxyl group on the same side of the polyketide chain. In addition, the KR has stereocontrol over the methyl group extending off the chain that results from the extension with methylmalonate. The KR is thus responsible for the orientation of two stereocenters. In some cases, such as the third module of DEBS, the KR is said to be "inactive," in which case the module contains a KR but does not reduce the carbonyl group; however, these domains still retain stereocontrol over the peripheral methyl group and thus are not inactive but rather represent a separate category of KR domains.

If the carbonyl group of the growing chain is reduced, the ACP may next pass to a dehydratase domain (DH), which catalyzes the dehydration of the hydroxyl formed at the KR, resulting in a C-C double bond. Again, the PKS exhibits stereocontrol, as a DH can catalyze the formation of the *cis* or *trans* isomer. Only the fourth module of DEBS contains a DH.

If a PKS module contains a full reductive loop, then the ACP will last transfer the growing chain to an enoyl reductase domain (ER). Here, the C-C double bond is reduced to completely saturate the original carbonyl carbon, as in fatty acid synthesis. These three domains – the KR, DH, and ER – are all optional, and the diversity of presence or lack thereof within the modules of a PKS results in the great diversity that we see in polyketides like 6-DEB.

After the ACP brings the polyketide chain through the reductive section of its module, it finally passes the chain to the KS of the next module, and the process is repeated. In this manner, DEBS begins with a starter unit of propionate bound to the KS of module 1 by the LDD and systematically adds six extender units of propionate, each modified by a subsequent module.

When the growing polyketide chain reaches the end of the assembly line, a terminal thioesterase (TE) domain catalyzes the hydrolysis of the polyketide from the phosphopantethienyl arm of the ACP to form a carboxylic acid, which, in the case of 6-DEB, cyclizes to form a macrolide. Thus, DEBS, the paradigmatic PKS, masterfully takes six molecules of methylmalonyl-CoA and one molecule of propionate-CoA and creates 6-DEB, the aglycone precursor to erythromycin.

2. Polyketide Libraries

Since each module of a PKS can perform different chemistry on the growing chain, the number of possible molecules that could be created by a PKS is seemingly endless. From **Figure 2** we see that three of the six DEBS modules – 2, 5, and 6 – perform the same chemistry on the growing polyketide, but even with only 4 distinct module types, there are 4^6 or 4096 different polyketides that could be produced from a six-membered PKS assembly line using DEBS modules. Include all of the different modules observed in nature from different organisms, and the number of possible polyketides is enormous.

By genetically engineering PKSs to contain novel sequences of modules, researchers are attempting to create exhaustive libraries of synthetic polyketides. These libraries are a common goal among polyketide scientists due to their potential as drug screening subjects and precursors for organic synthesis. Because synthetic polyketides share the structures of natural polyketides, many of which are now used as medicines, it is reasonable to hypothesize that libraries of polyketides will be better subjects for drug screening than other synthetic products. In addition, the presence of chiral groups makes polyketides attractive starting materials for the organic synthesis of "unnatural" natural products. Both of these applications give polyketides the potential to lead to the next generation of pharmaceutical drugs.

Standing in the way of polyketide libraries is the current protocol for purifying polyketides. Host cells engineered to produce the PKS of interest must be grown, and after they are induced to synthesize polyketides, the supernatant in which they grow must be extracted. The resulting product is very impure, and the yield of polyketide that results after column chromatography is extremely minimal – only a few milligrams per liter of cells. To bypass this obstacle, we designed a reduced-component *in vitro* system for the production of polyketides. We hypothesize that this system will dramatically increase the yield of polyketide production – making the goal of polyketide libraries a reality and unleashing the full potential of PKSs as manufacturers of new medicines.

B. *In Vitro* Polyketide Production

Before we could make polyketides *in vitro*, we first needed to design a suitable system. This system needed to contain the necessary components to perform three tasks. First, it needed to use enzymes to convert propionate into (2S)-methylmalonyl-CoA and propionyl-CoA, the substrates for DEBS (as described above). Next, the system needed to contain a PKS to take these substrates and convert them into the polyketide of interest. Finally,

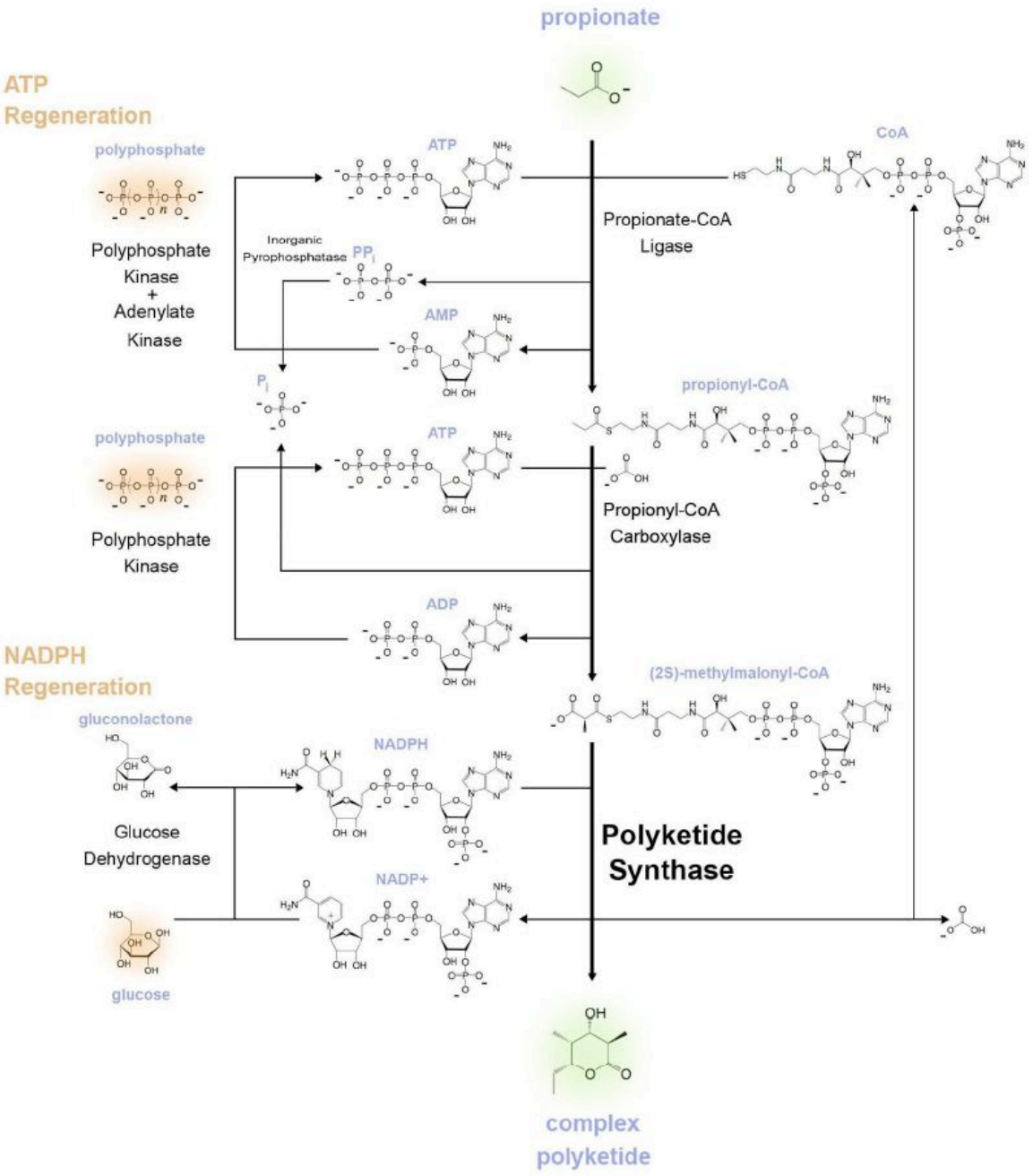

Figure 5: *In Vitro* System for Polyketide Production. The system includes enzymes necessary for methylmalonyl- and propionyl-CoA production as well as the regeneration of ATP and NADPH.

the system needed enzymes that could regenerate the energy-carrying cofactors necessary for polyketide production, ATP and NADPH, in order for polyketide yield to be substantial. In addition to these enzymes, we needed to include all the substrates necessary for these reactions and provide the ideal conditions for the enzymes with a buffer system. Once this system was constructed, the *in vitro* production of polyketides would be possible.

1. The Main Pathway

Before propionate can be incorporated into a polyketide by a PKS, it must first be modified. DEBS uses propionyl-CoA to prime its first module with propionate, as a FAS does with acetate. Each module then decarboxylates methylmalonyl-CoA in order to add a three-carbon extender unit to the growing polyketide chain. To form polyketides *in vitro*, we could simply add these substrates directly; however, it would be much cheaper to add propionic acid instead and form propionyl- and methylmalonyl-CoA *in situ*.

Streptomyces coelicolor bacteria contain genes for propionate-CoA ligase (PrpE) and propionyl-CoA carboxylase (a complex of two enzymes, AccA1 and PccB), which form propionyl-CoA and subsequently methylmalonyl-CoA, respectively. The first step of the *in vitro* system (**Figure 5**), then, involves these two enzymes, which transform cheap propionic acid into usable substrates for the PKS.

2. An Alternate Route

Another method to form propionyl- and methylmalonyl-CoA from simple substrates involves a different pair of enzymes. Erythromycin-producing bacteria *S. coelicolor* and *Saccharopolyspora erythrae* contain genes for methylmalonate-CoA ligase (MatB) and methylmalonyl-CoA epimerase. The first of these enzymes fuses coenzyme A to methylmalonate, forming a racemic mixture of methylmalonyl-CoA. Unfortunately, the (2R) epimer of methylmalonyl-CoA is a much more abundant product than the (2S) epimer, yet DEBS can only accept the latter form. In addition, the rate of spontaneous epimerization to the (2S) form is extremely slow (12). In order to produce the correct epimer, *S. coelicolor* and *S. erythrae* use methylmalonyl-CoA epimerase. Thus, an alternate pathway to the correct substrates for DEBS involves the addition of methylmalonate and these two enzymes.

3. Polyketide Synthase

Once the correct substrates are formed *in vitro*, a PKS must be present to transform them into the polyketide of interest. Once the *in vitro* system is completed and optimized, this element of the system will be exchanged for various PKSs, allowing for the production of polyketide libraries. In the mean time, we needed a simple PKS that would allow for the production of a familiar product that could be analyzed to quantify polyketide yield.

DEBS1 refers to the first of the three polypeptides of DEBS. It contains the first two modules of DEBS plus the LDD domain that primes DEBS with propionate. The addition of a three-carbon extender unit by each of the modules results in a triketide bound to the ACP of the second module. By engineering a TE domain at the end of module 2, this triketide can be cleaved off. In solution, this triketide will then cyclize to form a triketide lactone (TKL). We received plasmids containing the genes for the LDD-Mod1 of DEBS as well as Mod2-TE from Kosan Biosciences. After protein expression and purification, we inserted these proteins into the *in vitro* system to form the TKL shown in **figure 5**.

4. Energy Regeneration

The enzymes involved in polyketide production from propionate use the energy-carrying molecules ATP and NADPH. These substrates are expensive, so we included systems to regenerate these molecules in order to maximize the efficiency of the *in vitro* system.

Propionyl-CoA carboxylase uses ATP and bicarbonate to produce methylmalonyl-

CoA from propionyl-CoA, forming ADP in the process. In order to regenerate ATP from ADP, polyphosphate kinase (PPK) is used. This enzyme catalyzes the transfer of a phosphate group from polyphosphate, a very cheap substrate, to ADP to form ATP.

In the first step of the *in vitro* system, propionate-CoA ligase uses ATP to form propionyl-CoA, forming AMP in the process. To regenerate ATP from AMP, adenylate kinase (AdK) is first used to transfer a phosphate from ATP to AMP to form two molecules of ADP. Polyphosphate kinase can then regenerate ATP from ADP as discussed above.

Propionate-CoA ligase also forms inorganic pyrophosphate, the presence of which could lead to product inhibition. In order to rid the solution of pyrophosphate, inorganic pyrophosphatase (IP) is added to the system. This enzymes catalyzes the hydrolysis of inorganic pyrophosphate to form two molecules of phosphate.

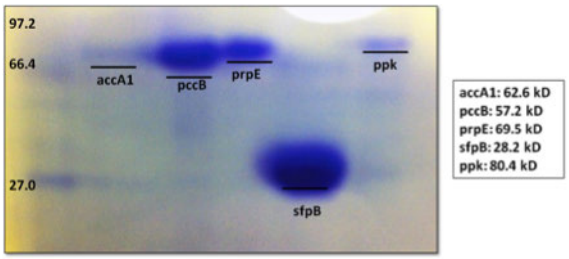

Figure 6: PAGE of several proteins in the *in vitro* system. Columns from left to right: protein ladder, propionyl-CoA carboxylase alpha subunit, propionyl-CoA carboxylase beta subunit, propionate-CoA ligase, Sfp (an enzyme for a separate project), and polyphosphate kinase.

Finally, the KRs of the PKS use NADPH to reduce the growing polyketide chain, forming NADP+ in the process. To regenerate NADPH, glucose dehydrogenase (GDH) is used to oxidize glucose, forming gluconolactone and transferring a hydride to NADP+. By regenerating the energy-carrying molecules of the *in vitro* system, the full potential of the system to increase the efficiency of polyketide production will be achieved.

5. Buffer Conditions

In addition to the enzymes described above, the correct buffer conditions and substrates must be added to the *in vitro* system. Imidazole at pH 7.8 provides a stable pH. Propionic acid provides the starting material for polyketide production. Energy-carrying molecules ATP and NADPH must be added, and $MgCl_2$ is added to stabilize ATP. $NaHCO_3$ is added for the carboxylation of propionyl-CoA. Polyphosphate allows for the regeneration of ATP as glucose does for NADPH. NaCl provides the correct salinity for the enzymes in the system, and BSA further stabilizes them. Finally, coenzyme A must be added to form propionyl- and methylmalonyl-CoA.

III. Results and Discussion

A. Protein Purification

The majority of the work thus far in the project has concerned the isolation of the enzymes in the system. While protein purification is usually a straightforward task, many of the proteins in the system have proven difficult to isolate. The genes for some required advanced PCR techniques to clone. In other cases, ligating the digested gene into an expression vector proved less than simple. Currently, we have isolated all of the enzymes in the original *in vitro* system and are assaying them for activity.

A few of the enzymes were purified with little trouble. Propionate-CoA ligase (PrpE), the beta subunit of propionyl-CoA carboxylase (PccB), and polyphosphate kinase (PPK) responded well to a standard PCR protocol, digestion and ligation procedure, and immobilized metal affinity chromatography (IMAC) (**figure 6**). Glucose dehydrogenase (GDH) and methylmalonate-CoA ligase were purified by other members of the lab. We received plasmids containing the genes for the LDD-Mod1 of DEBS as well as Mod2-TE from Kosan Biosciences. Plasmid vectors containing alternate versions of the propionyl-CoA

carboxylase enzyme complex proteins were also acquired from an outside laboratory and purified. The remaining enzymes were more difficult to isolate.

In order to clone the genes for the alpha subunit of propionyl-CoA ligase (AccA1), adenylate kinase (AdK), inorganic pyrophosphatase (IP), and methylmalonyl-CoA epimerase (MME), multiple PCR protocols were attempted simultaneously. Typically, nine protocols were attempted on each gene, utilizing a gradient of three DMSO concentrations and three different touchdown PCR protocols that utilized three different annealing temperature ranges. By performing this 3x3 PCR protocol experiment, we have found a PCR protocol that successfully clones every gene we have attempted to amplify.

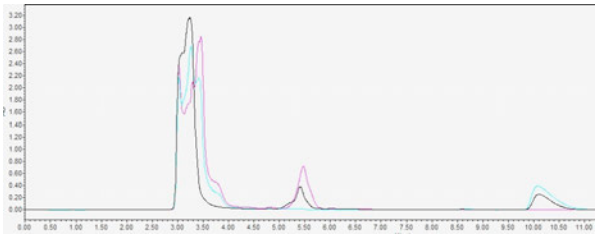

Figure 7: Ligase and Carboxylase Assay. The line drawn in pink represents the sample without enzyme. The blue line represents the addition of propionate-CoA ligase. The black line represents the addition of the propionyl-CoA carboxylase complex. See text for further description.

While most of the enzymes in the system have been successfully isolated, one still remains elusive. The gene for methylmalonyl-CoA epimerase has been successfully amplified from *S. erythrae* genomic DNA using the aforementioned PCR protocol, but several attempts at ligation into pET28b and transformation into Top Ten heat shock competent *E. Coli* have proven unsuccessful, as judged by check digest on the extracted DNA of the resulting colonies. Next, we will attempt to amplify this gene from *S. Coelicolor* instead, and hopefully the resulting product will allow us to successfully ligate this gene into pET28b.

B. Enzyme assays

Once each enzyme in the system was isolated, it was necessary to assay it for the correct activity. So far, all of the assays performed have utilized reversed-phase HPLC with UV-Vis spectrophotometric detection at 264 nm to measure the presence of adenine derivatives. To separate negatively charged adenine derivatives such as ATP, ADP, and propionyl-CoA on a reversed-phase HPLC column, we used a buffer containing tetrabutylammonium hydroxide (TBAH), whose butyl moieties bind to the hydrophobic column and ammonium moieties create anion-exchange conditions. The result is that more negatively charged species take longer to elute from the column.

1. Propionate-CoA Ligase

To assay PrpE, we combined ATP, propionate, and CoA with and without the enzyme and analyzed the samples with HPLC. **Figure 7** shows the disappearance of a peak for CoA at 5.5 mins and the appearance of a peak for propionyl-CoA at 10 mins when the enzyme is added, indicating that the desired reaction is taking place.

2. Propionyl-CoA Carboxylase

The propionyl-CoA carboxylase enzyme complex produces (2S)-methylmalonyl-CoA from propionyl-CoA and bicarbonate. To assay this enzyme, the product of the PrpE reaction was run on the HPLC with and without the addition of AccA1 and PccB. **Figure 7** shows the result of this addition. A peak with a retention time of about 5.5 minutes appears for the sample containing the carboxylase complex. Whether this peak represents methylmalonyl-CoA or unreacted CoA needs to be determined by running methylmalonyl-CoA alone on the HPLC.

3. Adenylate Kinase

AdK catalyzes the conversion of ATP and AMP to two molecules of ADP, which can then be regenerated to ATP by PPK. In order to assay AdK, we examined the progress of the reverse reaction, the conversion of ADP to AMP and ATP. **Figure 8** shows the affect of adding AdK to a solution of ATP. The resulting chromatogram shows three peaks when AdK is added, indicating the conversion of ADP to a mixture of ATP, ADP, and AMP.

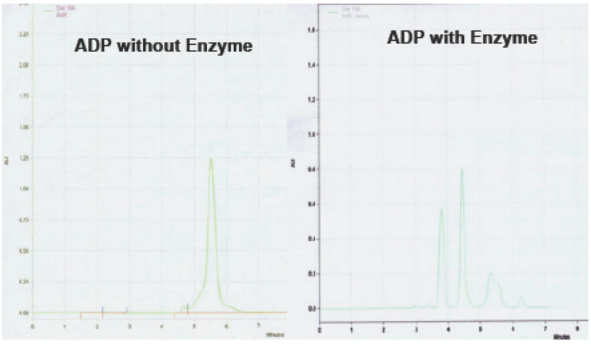

Figure 8: HPLC Assay of Adenylate Kinase. See text for more information.

2. Polyphosphate Kinase

The assay for PPK, which transfers a phosphoryl group from polyphosphate to ADP to regenerate ATP, is similar to the assays previously discussed. ADP was added in solution with polyphosphate, and this solution was run on the HPLC with and without the presence of PPK. The chromatogram showed a shift from ADP to ATP, indicating that the enzyme was regenerating ATP; however the reaction appeared to proceed at a slow rate.

3. *In Vitro* Polyketide Production

Once PrpE, AccA1, and PccB had been successfully purified, we combined these enzymes with LDD-Mod1, Mod2TE and the necessary cofactors required to produce a triketide lactone *in vitro*. To simplify the system at first, we used large concentrations of ATP and NADPH instead of adding the energy regeneration enzymes. The reaction was run overnight, extracted, and analyzed via LC-MS. Unfortunately, the triketide lactone was not detected. Before the complete system can produce a triketide lactone, we need to test all of the components for the correct activity, a process that is currently ongoing.

To demonstrate that polyketides can be made *in vitro* and to simultaneously test the activity of three enzymes, a series of reactions was performed that we hoped would result in the formation of a triketide lactone. First, MatB was incubated with methylmalonate to form methylmalonyl-CoA. Unfortunately, MatB produces only the (2R) epimer of methylmalonyl-CoA, the incorrect epimer for DEBS modules. To form a racemic mixture of methylmalonyl-CoA, the reaction was acidified with HCl and incubated overnight.

Meanwhile, the KR domain from the first module of DEBS, EryKR1, was incubated with propylpropyl-SNAC, a diketide, to produce a reduced diketide that DEBS Mod3 would accept.

The MatB reaction was incubated with the diketide reaction and Mod3-TE, another enzyme engineered by Kosan Biosciences, to hopefully produce a triketide lactone. This reaction was extracted and the resulting product was analyzed with LC-MS. Unfortunately, none of the mass spectra showed a peak for the mass of the triketide lactone, indicating that the reaction did not proceed as hoped. We will next analyze each step of this series of reactions to determine whether the correct intermediates are being formed.

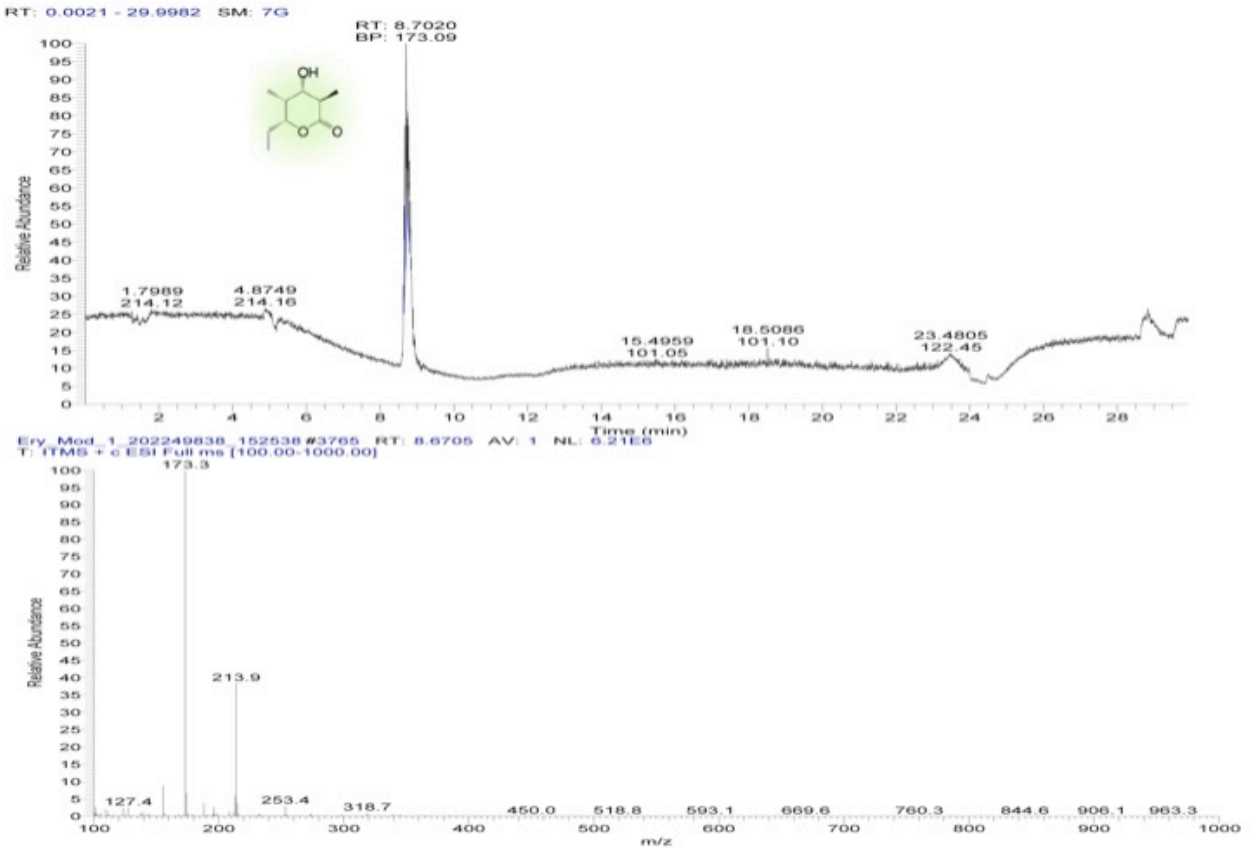

Figure 9: LCMS of triketide lactone produced using the current method.

C. Positive control

We used the current polyketide extraction protocol to produce the triketide lactone that we hope to synthesize with the *in vitro* system as a positive control. K207-3 *E. Coli* engineered by Kosan Biosciences to produce the triketide lactone in question were grown, and their media was extracted. The resulting product was purified with silica column chromatography, and the sample was analyzed with LCMS (**figure 9**).

IV. Conclusion

So far, we have made great progress toward our goal of constructing an *in vitro* system for the preparative biosynthesis of polyketides. We have successfully purified most of the enzymes in the system, and unique assays have been performed on each enzyme isolated to test its activity. Unfortunately, despite this progress, our first attempts at putting the system together have failed to produce a measurable amount of polyketide as judged by LCMS.

The lack of polyketide production in these preliminary attempts likely stems from the absence of activity of one of the enzymes in the main pathway. AccA1, the alpha subunit of propionyl-CoA carboxylase, proved difficult to isolate at first, and even after it was successfully purified, our attempts to assay the propionyl-CoA carboxylase complex proved inconclusive due to inconsistencies in our HPLC assay results. We have obtained alternate versions of the enzymes in the complex in hopes that these would give more consistent results. Meanwhile, we are developing different

HPLC methods to assay this complex that will hopefully be more reliable.

Concurrently, we have attempted to explore an alternate pathway to methylmalonyl-CoA that would replace propionate-CoA ligase and propionyl-CoA carboxylase. Instead, methylmalonate would be added directly in solution, bound to coenzyme A by methylmalonate-CoA ligase, and epimerized to the (2S) form with methylmalonyl-CoA epimerase. Unfortunately, our attempts to clone the gene for the latter enzyme have been thus far unsuccessful. We are currently troubleshooting this purification process, and hopefully we will be successful in our attempts soon.

We have also begun investigating a more sensitive method for detecting polyketide production. If C^{14}-labeled methylmalonate or propionate is used in the system, then radiolabeled polyketides should be formed. These polyketides can be quantitatively detected by using thin layer chromatography and a radioisotope detection system. This capability will allow us to more accurately determine the amount of polyketide being produced with the *in vitro* system, and it will facilitate our optimization of the system once we are successful at producing polyketides.

While our attempts at producing polyketides *in vitro* have been unsuccessful so far, we are making progress toward this goal and are optimistic. Once we are successful at synthesizing polyketides with the *in vitro* system, we hope that the system will greatly increase the cost and time efficiency of polyketide production, increasing the yield dramatically over that of the current method. We hope that this increase in efficiency will allow us to realize our long-term goal of creating exhaustive polyketide libraries that will lead to a generation of new molecules and new medicines.

V. Acknowledgements

The Welch Foundation has sponsored this project. Engineered polyketide synthases were obtained from Kosan Biosciences. Plasmids for the subunits of propionyl-CoA carboxylase were obtained from Sheryll Tsai of UC Irvine.

Shawn Piasecki purified several of the enzymes in the system. Andrew Harper was responsible for many of the HPLC assay chromatograms. Amanda Hughes successfully purified MatB. Jianting Zheng, Clint Taylor, and June Liu provided help with laboratory methods. Dr. Adrian Keatinge-Clay supervised this project and provided guidance.

VI. Materials and Methods

The following section serves as a reference to those wishing to replicate the procedures performed over the course of this project. Herein are listed the various protocols used to clone and purify each protein, conditions for each assay performed, and a detailed description of the current method for polyketide production.

A. Protein Cloning Protocols

1. Standard Cloning Protocol

Several of the proteins purified were isolated using the following protocol: A standard PCR recipe was performed, and the following temperature conditions were run on a thermal cycler: 95˚C for 3 min, followed by 34 cycles of 95˚C for 30 sec, 65˚C for 30 sec, and 72˚C for 30 sec, followed by 72˚C for 10 min.

The resulting fragment was digested with Nde1 and Xho1 restriction endonucleases and ligated into pET28b. The recombinant vector was transformed into electrocompetent XL1-blue cells via electroporation and spread on a kanamycin-resistance selection plate of LB/agarose. The resulting colonies were grown in 3ml of LB media supplemented with kanamycin, and the plasmid DNA was subsequently extracted and transformed into K207-3 cells via electroporation. These cells were grown in 6L of LB media supplemented with kanamycin to an OD_{600} of 0.5 and induced with .5mM IPTG. The cells were then grown overnight at 15˚C.

These cells were centrifuged at 5,000 x g for 15 mins, and the resulting cell pellet was sonicated in a lysis buffer consisting of .5M NaCl and 30mM HEPES pH 7.5 to release the protein into solution. The lysate was centrifuged at 30,000 x g. The supernatant of this centrifugation was run over an immobilized nickel affinity column primed with the aforementioned lysis buffer. The column was washed with 15mM imidazole in lysis buffer, and the protein was eluted with 150mM imidazole in lysis buffer.

Using this protocol, adenylate kinase was cloned from the genome of K207-3 *E. coli* using the following PCR primers:
F: 5' TCAGTAAGTCATATGTTATTTATTCTTTGCGCGCTC
R: 5' AGAGTAGACCTCGAGATGAGCTTACTCAACGTCCCTG.

Propionate-CoA ligase was cloned from the genome of *S. coelicolor* using the following PCR primers:
F: 5' GATGTTGCACATATGTCTTTTAGCGAATTTTATC
R: 3' GCGGTCCGCTACCTTCTCATTGAGCTCCAAGCATAG.

The beta subunit of propionyl-CoA carboxylase was cloned from the genome of *S.coelicolor* using the following PCR primers:
F: 5' GATCTGTATCATATGTCCGAGCCGGAAGAGCAG
R: 3' GTGCCGTTGTAGGGGGACATTGAGCTCTAGCATTCA.

Polyphosphate Kinase was cloned from the genome of *S. coelicolor* using the following PCR primers:
F: 5' GATACTGTACATATGGGTCAGGAAAAGCTATA
R: 3' AGTGAGCTTGTTGGACTTATTGAGCTCTAGCACTGA.

2. Inorganic Pyrophosphatase

Inorganic pyrophosphatase was purified in a manner identical to that above except that the PCR protocol was modified to have an annealing temperature of 79°C, and 10% dimethyl sulfoxide was added to the PCR reaction. The following primers were used:
F: 5' TCAGTAAGTCATATGTTATTATTCTTTGCGCGCTC
R: 5' AGAGTAGACCTCGAGAGAGCTTACTCAACGTCCCTG. All other procedures were the same as those listed above.

3. Propionyl-CoA Carboxylase Alpha Subunit

The procedure above was modified significantly to purify AccA1. A touchdown PCR protocol was utilized along with 10% dimethyl sulfoxide. The touchdown protocol temperatures were as follows: 95°C for 3 min, followed by 5 rounds of 95°C for 30 sec, 66°C for 30 sec (2°C cooler at each step), and 72°C for 1 min. This was followed by 30 rounds of 95°C for 30 sec, 60°C for 30 sec, and 72°C for 1 min. The protocol ended at 72°C for 8 min. The protein was cloned from the genome of K207-3 *E. coli* using the following PCR primers:
F: 5' GATGTTCAACATATGCGCAAGGTGCTCATCGCC
R: 3' GTAGACGCTCTAGTTCCTGACTCTAAGTAGCAATCG. Nde1 and EcoR1 endonucleases were used. All other procedures were the same as above.

4. Propionyl-CoA Carboxylase from *Mycobacterium tuberculosis*

Plasmids containing genes for the alpha and beta subunits of propionyl-CoA carboxylase were obtained from an outside laboratory. These plasmids were transformed into K207-3 *E. coli* by electroporation and grown as above to an OD_{600} of .6. For the alpha subunit, the cells were then induced with .5mM IPTG and grown at 20°C overnight. For the beta subunit, the cells were induced with .1mM IPTG and grown at 18°C overnight. The lysis, wash, and elution buffers used for protein purification contained 10% w/v glycerol, 150mM $NaPO_4$ and .5M NaCl.

B. Detailed Assay info

1. Adenylate Kinase

A solution of 150mM NaCl, 100mM HEPES pH 7.5, 1mM ADP, and 10mM $MgCl_2$ was prepared with and without the addition of adenylate kinase. The resulting two reactions were run on a reversed-phase HPLC with an isocratic elution system of methanol supplemented with tetrabutylammonium hydroxide. The separation was analyzed with UV spectroscopy at 264 nm.

2. Propionate-CoA Ligase and Propionyl-CoA Carboxylase

A solution of 100mM KH_2PO_4, 12mM $MgCl_2$, 50mM $NaHCO_3$, 20mM propionic acid, 5 mM ATP, and 1mM coenzyme A was prepared with and without the addition of propionate-CoA ligase. These reactions were run as above via HPLC. The subunits of the propionyl-CoA carboxylase complex were then added to the solution containing the ligase, and this reaction was run on the HPLC as well.

3. Methylmalonate-CoA ligase

A solution of 100mM KH_2PO_4, 12mM $MgCl_2$, 10mM methylmalonic acid, 5 mM ATP, and 1mM coenzyme A was prepared with and without the addition of methylmalonate-CoA ligase. These reactions were run as above on the HPLC. Data for this assay has not yet been collected.

C. Detailed current method protocol

In a large Fernbach flask, we combined M9 media, $CaCl_2$, $MgSO_4$, glucose, β-alanine, ampicillin, kanamycin, and *E. coli* cells engineered previously to produce the desired triketide lactone. These cells were grown at 37°C until they reached an OD_{600} of 0.4. At this point, the heat was reduced to 22°C, and propionic, succinic, and glutamic acids were added. The cells were incubated for 15 min before IPTG was added to induce polyketide production. The cells were then incubated overnight.

After incubation, the cells were centrifuged at 6000 RPM for 10 min. The supernatant was again centrifuged at 7500 RPM. The supernatant from this centrifugation was pHed to 1.5 with approximately 20ml concentrated HCl. The supernatant was then extracted twice with ethyl acetate, and $MgSO_4$ was added to desiccate the solution. The organic layer was removed, and the solvent was evaporated off. The resulting product was analyzed via a vanillin-stained TLC and LCMS. The vanillin stain consisted of 6g vanillin, 100ml 95% ethanol, and 1ml concentrated H_2SO_4.

VII. References:

1. Hopwood, D.A. (1997) *Chem Rev*, 97, 2465–2498.
2. Hertweck, C. (2009) *Angew Chem Int Ed,* 48, 4688 – 4716.
3. Hagan, D. O. The Polyketide Metabolites, Ellis Horwood, Chichester, 1991.
4. Khosla, C., et al. (2007) *Annu Rev Biochem,* 76, 195-221.
5. Gokhale, R. S., et al. (1999) *Science,* 284 (5413), 482-485.
6. Khosla, C., et al. (1999) *Annu Rev Biochem,* 68, 219-253.
7. McDaniel, R., et al. (1999) *Proc Natl Acad Sci,* 96, 1846-1851.
8. Hilterhaus, L., Liese, A. (2007) *Adv Biochem Eng Biotechnol,* 105, 133-173.
9. Watkins, V. S., et al. (1997) *Annals of Pharmacoterapy,* 31(3), 349-356.
10. Cane, D. E., et al. (1998) *Science,* 282 (5368), 63-68.
11. Kumar, P., et al. (2003) *J Am Chem Soc,* 125(14), 4097-4102.
12. Fuller, J. Q., Leadlay, P. F. (1983) *Biochem J*, 213, 643-650